

ACTIVE MONSOON AND DAILY RAINFALL IN THE HILL SUB REGION OF WEST BENGAL – AN INFORMATION THEORY APPROACH

SATHISH. G & BANJUL BHATTACHARYYA

Department of Agricultural Statistics, Bidhan Chandra KrishiViswavidyalaya, Mohanpur, Nadia, West Bengal, India

ABSTRACT

The probabilities of occurrence of dry and wet days is one of the best tools for fitting a Markov chain model, geometric and Markovian geometric spell distribution. But, the studies of uncertainty about the nature of day's rainfall are rarely available. Every probability distribution has some uncertainty associated with it. The concept of Information theory approach (entropy) is introduced here to provide a quantitative measure of this uncertainty, and to test the stochastic Markovian dependence. Present study reveals that redundancy values (R) in Hill sub region of West Bengal i.e., Darjeeling, during monsoon months are ranging from 0.116 to 0.536, these values are indicating that Markovian system tends to minimum favorable conditions i.e., almost uncertain to follow Markovian dependence. But, likelihood ratio test is the most powerful test to check the Markovian dependence. Using likelihood ratio test, Markovian dependence is best fitted during the study period, except in September.

KEYWORDS: Darjeeling, Entropy, Information Theory, Markovian Dependence, Probability Distribution & Redundancy Test

Received: Jul 10, 2017; **Accepted:** Jul 27, 2017; **Published:** Jul 29, 2017; **Paper Id.:** IJASRAUG201764

INTRODUCTION

Rainfall is considered as a principal source of water. The success or failure of crops, particularly under rainfed condition is closely linked with rainfall distribution patterns. The seasonal, spatial and temporal variations of rainfall adversely affect the crop production under dry land / rainfed situation. The crop cultivation under dry land situation involves high risks of crop failure; therefore, the farmers are not willing to make high investment. Under such situation, analysis of rainfall distribution pattern is very much essential to utilize maximum potential of dry land agriculture. About one-third of total cultivated land (140 million ha) is irrigated, producing 55 percent of food grain production and remaining cultivated area, which is mainly dry land / rain fed, which contributes 45 percent. The net sown area of about 140million ha provides food to over one billion population (Shamsher Singh, 2006).

In West Bengal, crop planning is done, based on the availability of irrigation. But, it is restricted due to the variability in the beginning of the monsoon season. The agricultural production in the rain fed areas of West Bengal suffers mostly from a late start of the monsoon. Also, if the withdrawal of monsoon from the region is earlier than expected, late planted crops may be hurt during the filling stages, from lack of moisture. The systematic studies on the influence of rainfall on the yield of wheat at Rothamsted Agricultural Experiment Station was first done by Fisher (1924), and reported that the yield of crop was more affected by distribution of rainfall than the amount. Studies on the probability of occurrences of dry and wet days using Markovian model was done by Gabriel and Neumann (1962) and Medhi (1976). Markovian model can be used to study the probability of

occurrences of different rainfall characteristics through one step transitional probability matrices during monsoon period, and revealed that, the weighted entropy values over the transition probabilities during the beginning and closing months of monsoon at Maithan were less, in comparison with the other monsoon months. Basu, 1988 and Thokalet *al.* (2003) studied the rainfall data of 31 years (1972 – 2002) of Dapoli, located in Konakan region of Maharashtra state. The statistically analyzed weekly data of rainfall was found to be more useful for crop planning as water management practices, than monthly, seasonal and annual data. Maruyama and Kawachi (1998) and Kawachi et al. (2001) employed the informational entropy to measure the degree of uncertainty of rainfall occurrence in time or the over-a-year temporal rainfall apportionment based on the probability density function of rainfall, randomly apportioned over fragmented times. Pechlivanidis et al. (2010) used Entropy for Model diagnostic in rainfall modeling. Saxena et al. (2013) studied the order identification of daily rainfall by using Akaike's criterion. But, Shannon's entropy has not been yet used to study the uncertainties of rainfall and an attempt has been made in this paper, to analyze the daily rainfall data of Hisar station for the period from 1925 to 2013, using Shannon's entropy. Mukhopadhyay (1993) attempted to examine the Markov dependence by entropy, considering the three states of outcome. There were normal, bad and good years of rainfall, accordingly defined. Redundancy test was applied to test the stochastic dependence on the one step 3x3 Markov chain model. But it was observed from the study that, the redundancy test could not favor the system of Markovian dependence. Mukhopadhyay (1996) attempted to use entropy for Markovian model on weekly rainfall in two North Bengal districts. A statistical study of monsoon rainfall at two stations of Terai zone during 1972- 1990 was carried out by Markov chain model through one step transition probability matrix. The stochastic matrix for the two stations was obtained from one transition state to others for different nature of weekly rainfall. The likelihood ratio test based on entropy was applied over Markovian model. It was found that the weekly rainfall during rainy months depend on the nature of rainfall of the previous week. Markov chain models were fitted to weekly rainfall during monsoon season for the two stations, Cochin and Jalpaiguri, but the stochastic matrices of the two places were not considered to be identical.

Keeping in view of the above consideration, it is utmost importance to study the Information theory approach regarding uncertainty of monsoon rainfall, in the hill region of West Bengal.

MATERIALS AND METHODS

Forty-four years of daily rainfall data of active monsoon period (from May 31st to October 30th) of meteorological station - Darjeeling was used in this study. The study was considered from 1970 to 2014. To study the probabilities of dry and wet spell and uncertainty of active monsoon rainfall, the daily rainfall was collected from Indian Meteorological Department (IMD).

Stochastic Matrices

A unit step 5x5 transition probability matrix, from one transition state to another was formed for each month, June to October, and the season from the frequency of the occurrence of daily rainfall at Darjeeling meteorological station.

In the present study, the choice of minimum threshold values for determining the states of Markovian model on daily rainfall are of much implications, whenever this information of a region are extremely valuable in planning agricultural operational adjustments in rain fed farming, and also critical to plants as well as various human activities. Bearing it in mind, the intensity of daily rainfall have been assigned to five categories and grouped as follows:

Table 1

Category No.	Nature of Daily Rainfall	Intensity of Daily Rainfall (in mm)
A	Non rainy day	0-2.4
B	Light rainy day	2.5-10.0
C	Moderate rainy day	10.1-50.0
D	Heavy rainy day	50.1-125.0
E	Very heavy rainy day	125.1 and above

The mode of classification yields a sequence of non-rainy, light rainy, moderate rainy, heavy rainy day, very heavy rainy days which can be regarded as a five states Markov chain. The classification was successfully used by Basu (1988). There are twenty-five possibilities in which, each daily rainfall with its transition may be classified as one of them. Thus for each year, the nature of rainfall on 1 st June was classified as one of them, which also depends on the nature of rainfall on 31st May for transition property. Repeating this procedure for each year under the experimental period, the cell frequencies for twenty-five possibilities were obtained, accordingly. The nature of the intensity of rainfall on 31st May was regarded as the initial state of the chain, under study. Let the cell frequency of unit step transition $i \rightarrow j$. i.e. the process can go from i to j th state of occurrence and is denoted by n_{ij} , where i and j belongs to S ; $S = 1, 2, \dots, m$ (5) and also S is a closed set of five states. The set of m^2 (25) transition frequencies have sufficient statistics for the transition matrix. The transition probabilities have been estimated by applying the maximum likelihood method.

The estimate of the conditional probability is given by

$$p_{ij} = p[x_n = j / x_{n-1}] \quad (1.1)$$

Where x_{n+1} be the realization and i and $j \in S : S = 1, \dots, m$.

The initial probability

$$p_u^0 = P[x_0 = u] : u \in S.$$

But, it has been ignored from the chain. The transition probabilities are calculated by dividing the transition frequencies by the total frequencies of that particular state, and is given by

$$p_{ij} = \frac{n_{ij}}{n_i} \quad (1.2)$$

Where $n_i = \sum_j n_{ij}$. Subject to the restrictions: $\sum_j p_{ij} = 1$ for each i

And $P_{ij} \geq 0$ for all i and j

Each row constitutes the probability vector, and it is convenient to give the set of all such vectors as a matrix. This matrix is known as transition probability matrix or stochastic matrix. This is a square matrix (5x5) with non-negative elements and unit row sum.

The same method of estimation has also been employed to estimate the stationary probability of each state and is given by

$$\pi_i = \frac{n_i}{n} \quad (1.3)$$

Where $n = \sum_i n_i$

Subject to condition $\sum_i \pi_i = 1 \text{ } i \in S.$

The method of estimation of transition probability and stationary probability has been discussed by Bhattacharyya and Waymire (1990) and Seneta (1981). The values of stationary probability for each month and season are given in table 1 for Darjeeling respectively.

Redundancy Test

The mathematical model of information theoretic entropy given by Shannon (1948) has been employed to measure the uncertainty of the transition probability and is given by

$$H_i = - \sum p_{ij} \log p_{ij} \text{ for each } i \quad (1.4)$$

Where, H_i denotes the entropy of i th state. The entropy for each state of transition matrix is to be measured separately for the month of June, July, August, September and October, as well as for the monsoon season at all the stations. These are presented in table 1 for Darjeeling, respectively.

The entropy of stationary distribution i.e. the entropy of the individual state of occurrence is given by

$$H_{11} = - \sum_i \pi_i \log \pi_i \quad (1.5)$$

Where, π_i is the unconditional probability of the i^{th} state of occurrence, as given in table 2. The values of the entropy of the individual state of occurrence have been given in table 3. The entropy has an important use in measuring the uncertainty as well as in testing the hypothesis of Markov dependence. The calculation of the second entropy is based upon the weighted average of the entropies for each transition state. The weighted entropy is given by

$$H = - \sum_i \pi_i H_i \quad (1.6)$$

Where, π_i is the probability of individual state of occurrence. i.e the i th state and H_i is the entropy of the i th state. The values of the weighted entropy also presented in table3, Darjeeling, respectively.

A sequence of uncertainty, M , of the stationary model is obtained from the individual states of occurrence over the Markovian model in the system and is given by

$$M = H_{11} - H = (- \sum_i \pi_i \log \pi_i) - \sum_i \pi_i H_i \quad (1.7)$$

The redundancy of the state of occurrence, R , is obtained as the difference from one of the ratio of the weighted entropy value H to the maximum possible entropy (H_{\max}).

Here, $H_{\max} = \log 5$, as we have only five state of occurrence as Thail (1973).

So,

$$R = 1 - \frac{H}{H_{\max}} \quad (1.8)$$

This redundancy value is used to determine the favorableness and unfavorableness of the Markovian system. As the redundancy value, R , tends to be 1, the Markovian system tends to maximum favorable condition i.e. almost certain. Now on the light of this argument, we may examine the Markovian dependency on the monsoon months and season.

Likelihood Ratio Test by Entropy

In this section, we use the informational measure to test the hypothesis of Markovian dependence. The mathematical model of Shannon is used to obtain the measure of entropy in individual state of occurrence, denoted by H_{11} . The average conditional uncertainty can be measured by

$$H_{22} = - \sum_i \pi_i P_{ij} \log P_{ij}. \quad (1.9)$$

This is the same as the weighted entropy.

The hypothesis testing, involving Markov chain has been considered by several ways. Mainly the Chi-square and the likelihood ratio criterion have been used for testing the hypothesis of independence of the random variable. Here, we introduce a test criterion, which involves the entropy but, this is equivalent test of the likelihood ratio criterion. The test statistic is

$$\begin{aligned} T_1 &= 2n (H_{11} - H_{21}) \\ &= 2 \sum n_{ij} \log \frac{n_{ij}}{n_i n_j} \end{aligned} \quad (1.10)$$

That is, T_1 is the same as the likelihood ratio criterion. The test statistic has a limiting Chi-square distribution with $(m-1)^2$ degree of freedom and the large values of the statistic correspond to rejection of the hypothesis.

RESULTS AND DISCUSSIONS

Transitional probability matrix of 5x5 state was calculated, results are given in table 1. In Darjeeling region, non-rainy day followed by non-rainy day is the highest in the month of October (.871); however over the season it is 0.077. Light rainy day followed by light rainy day is the highest in the month of October (0.316), over the season, it is 0.086. Moderate rainy day followed by moderate rainy day highest is in July (0.445), and over the season it is 0.367. Heavy rainy day followed by heavy rainy day is highest in the month of June (0.409), and over the season it is 0.459. Very heavy rainy day followed by very heavy rainy day also highest in the month of July (0.146) and over the season 0.369.

Using transitional probability matrix, unconditional probabilities was calculated. From the table 2, it is found that in Darjeeling, the unconditional probability of non-rainy day is highest, followed by moderate rainy day during the monsoon month and season, except the month of July and October, where the highest is moderate rainy day followed by non-rainy day in July, and light rainy day followed by moderate rainy day in October.

Next step is to calculate entropy values, using conditional and unconditional probability matrix, From the table 3 it is found that during the monsoon season (June – October), the degree of uncertainty of very high rainy day is more than

the other state at Darjeeling (0.65) in August, followed by high rain day (0.64) in June. Considering the values of the entropy of the stationary distribution of study meteorological observatory, it is necessary to regard the occurrence of rainfall at Darjeeling is in highest uncertainty. The weighted entropy values of the conditional probabilities during the monsoon period is the highest uncertain in the months of June and July, followed by August and September. But, considering the empirical result of Medhi (1976), it would have been suggested that the likelihood ratio test is more powerful as well as more appropriate test against the Markov dependence. So, it is necessary to test Markovian dependence, using likelihood ratio test. From the table 4.4 it is observed that, the calculated value of likelihood ratio test by entropy (T_1) is significant at both 5% and 1% level of significance, except in the month of September. Thus, it indicates rejection of null hypothesis of independence against Markovian dependence. Therefore, at the statistical point of view, Markov Chain models have been fitted significantly on the occurrence of daily rainfall with its intensity, which might depend on the state of occurrence of rainfall on the previous day only.

Table 1: Transition Probability Matrix for Darjeeling during Monsoon Months and Season

JUNE						JULY					
	A	B	C	D	E		A	B	C	D	E
A	0.471	0.171	0.267	0.076	0.016	A	0.382	0.216	0.307	0.088	0.007
B	0.300	0.261	0.332	0.091	0.016	B	0.211	0.299	0.316	0.163	0.010
C	0.268	0.197	0.362	0.145	0.027	C	0.169	0.189	0.445	0.161	0.035
D	0.172	0.088	0.302	0.409	0.028	D	0.116	0.152	0.438	0.246	0.049
E	0.172	0.069	0.552	0.207	0.000	E	0.098	0.146	0.415	0.195	0.146
AUGUST						SEPTEMBER					
	A	B	C	D	E		A	B	C	D	E
A	0.486	0.185	0.265	0.051	0.013	A	0.629	0.129	0.177	0.062	0.003
B	0.289	0.315	0.305	0.074	0.017	B	0.351	0.243	0.315	0.068	0.023
C	0.254	0.182	0.417	0.120	0.028	C	0.319	0.165	0.359	0.134	0.022
D	0.238	0.206	0.365	0.159	0.032	D	0.343	0.153	0.270	0.212	0.022
E	0.226	0.194	0.355	0.129	0.097	E	0.333	0.167	0.278	0.222	0.000
OCTOBER						JUNE-OCTOBER					
	A	B	C	D	E		A	B	C	D	E
A	0.871	0.057	0.058	0.011	0.004	A	0.077	0.077	0.462	0.308	0.077
B	0.496	0.316	0.120	0.060	0.008	B	0.000	0.086	0.429	0.343	0.143
C	0.500	0.167	0.242	0.083	0.008	C	0.019	0.028	0.367	0.428	0.157
D	0.471	0.147	0.235	0.059	0.088	D	0.007	0.020	0.239	0.459	0.276
E	0.100	0.300	0.300	0.200	0.100	E	0.003	0.024	0.168	0.436	0.369

Table 2: Stationary Probability for Darjeeling of Monsoon Months and Season

	June	July	Aug	Sep	Oct	Season
pi1	0.323	0.212	0.338	0.473	0.788	0.009
pi2	0.181	0.211	0.214	0.159	0.094	0.025
pi3	0.321	0.387	0.336	0.256	0.086	0.261
pi4	0.155	0.161	0.090	0.098	0.024	0.440
pi5	0.021	0.029	0.022	0.013	0.007	0.265

Table 3: Information in Bits for Darjeeling during Monsoon Month and Season

	June	July	Aug	Sep	Oct	Season
H1	0.551	0.568	0.531	0.457	0.225	0.570
H2	0.591	0.606	0.584	0.583	0.509	0.529
H3	0.616	0.603	0.598	0.601	0.537	0.520

Table 3: Contd.,						
H4	0.584	0.604	0.624	0.617	0.590	0.506
H5	0.496	0.640	0.657	0.588	0.654	0.493
H11	0.612	0.618	0.593	0.556	0.325	0.521
H22	0.583	0.593	0.579	0.554	0.442	0.535
M	0.029	0.024	0.014	0.002	0.117	0.014
R	0.125	0.116	0.152	0.205	0.536	0.255

Table 4: Likelihood Ratio Test Statistic (T_1) Values

Station	June	July	Aug	Sep	Oct	Season
Darjeeling	77.33	68.02	38.21	5.53	327.32	39.01

* Degree of freedom is 16

CONCLUSIONS

To study the rainfall distribution and its characteristics, Markov chain model is one of the best tools that are available. Also, it is important to study the uncertainty about rainfall. Using redundancy test to estimate entropy values for monsoon rainfall in hill sub region of West Bengal, uncertainty is calculated. From results, redundancy test used to estimate entropy values, which ranged from 0.116 to 0.536 shows that monsoon rainfall in hill sub region of West Bengal, is highly uncertain. Likelihood ratio test is more powerful as well as more appropriate in testing the Markov dependence. Except in the month of September, the results during this study period from June to October showed that, likelihood ratio test are more significant, depicting that system to follow mark vain dependence.

REFERENCES

1. Basu, G.C. (1988). A study of monsoon daily rainfall at maithan by markovian model and information theory. *Mausam* 39, 83-86.
2. Bhattacharyay, R.N and Waymire, E.C(1990). *Stochastic Process with Application*. John Wiley and Sons, New York.
3. Fisher, R. A. (1924). *Phil. Trans. B*, 213, 89
4. Gabriel, K.R and J. Neumann. (1962). A markov chain model for daily rainfall occurrence at Tel Aviv. *Quart, J.R. Met. Soc.* 88, 90-95.
5. Kawachi, T., Maruyama, T., Singh, V.P., (2001). Rainfall entropy for delineation of water resources zones in Japan. *Journal of Hydrology* 246, 36–44.
6. Maruyama, T., Kawachi, T., (1998). Evaluation of rainfall characteristics using entropy. *Journal of Rainwater Catchment systems* 4 (1), 7–10.
7. Medhi, J (1996). *Stochastic Processes*. New age International (P) Ltd. Calcutta.
8. Mukhopadhyay, S. K. (1993). On the use of entropy in Markovian model of rainfall. *Mausam*, 45, 75-78.
9. Das, A. D and Mukhopadhyay, S. K (1996). A study on effective monsoon and dry spells of short return periods during monsoon months in two north Bengal districts. *Mausam*, 47, 2, 145-148.
10. Penchivanidis, I.G., Jackson, B, and McMillan, H. (2010): In the proceedings of International Environmental Modelling and Software Society.
11. Saxena, K.K., ruhal, Lokesh and Rao,V.U.M. (2013): Akaike Information Criterion for order identification of Daily rainfall at Hisar. *International Jr. Agricult.Stat.Sci.* Vol 9,I.71-78.

12. Seneta, E (1971). *Non- negative matrices and markov chains*. Springer- Verlog. New York.
13. Shamsher Singh.,(2006) *Natural resources management for sustainable Agriculture production- status and strategies*. *Journal of Soil Water Conservation*. 5 (3) : 16 – 28
14. Shannon, Claude E. (July–October 1948). "A Mathematical Theory of Communication". *Bell System Technical Journal*. 27 (3): 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x
15. Thokal, R.T., Mahale, D. Shetye, V.A. Pimple, D.P. and Powar, A.G. (2003). *Rainfall probability analysis for crop planning at Dapoli*. In: *Integrated water resources planning and management*. pp 51-58.